Achievement of Market-Friendly Initiatives and Results Program
(AMIR 2.0 Program)

Funded by U.S. Agency for International Development

Introduction to Basic Concepts and Comments on Learning Outcomes (CFA I, 2003)

By
Ronald E. Copley, ph.D., CFA
President
Copley Investment Management

Contract No. 278-C-00-02-00210-00

2003

Quantitative Analysis

Introduction to Basic Concepts and Comments on Learning Outcomes (CFA I, 2003)

Note to candidates: You will need to refer to the Learning Outcomes in your Study Guide for wording of each LO. I have provided you with comments on each LO in each study session, but have not duplicated the LOs in these notes.

by

Ronald E. Copley, Ph.D., C.F.A.
President
Copley Investment Management
5218 Wrightsville Ave.
Wilmington, NC 28403 USA
T: 910-452-7147
F: 910-395-6849
E: RonCopley@AOL.com

Table of Contents

Introductory Comments

Frequency Distributions and Graphic Presentation

A frequency distribution groups data into nonoverlapping categories and shows the number of observations in each category. The following steps are used to develop a frequency distribution: Determine the size of the class interval. If the number of classes has been determined, the following formula will provide a guide to the size of the class interval:

(Highest value - Lowest value) / Number of classes

If the number of class intervals is unknown the following formula will provide a guide to the size of the class interval:

(Highest value - Lowest value) / (1 + 3.222*Logarithm of the total frequencies)

Another method of determining the suggested number of classes is to find the smallest integer k such that 2^k => n, where n is the number of observations.

Tally the raw data into the classes to arrive at the frequency distribution.

There are four rules to constructing a frequency distribution:

1) Avoid having fewer than 5 or more than 15 classes.
2) Avoid open-end classes
3) Keep class-intervals the same size
4) Do not create overlapping classes

There are several type of graphs used to present frequency distributions:

- Histogram shows the number of frequencies in each class in the form of bars,
- A frequency polygon and a relative frequency polygon have the classes scaled on the x-axis and the class frequencies on the y-axis. Class midpoints and the corresponding class frequency are plotted as a point. These points are connected to form the polygon. The area under the polygon is equal to the number of frequencies.

A Relative frequency distribution is similar to the frequency distribution; the difference is it presents the percent of total observations in each class. Stem and leaf displays organize the raw data into some meaningful form. Data are partitioned into a stem and a leaf. The first digit of the number is the stem and the trailing digit is the leaf.

Cumulative frequency polygons are of several types:

- A less than cumulative frequency polygon allows the user to determine how many, or what percent of the observations are equal to or less than a certain value.
- A more than cumulative frequency polygon is constructed by adding the class frequencies starting with the highest class. Each of the lower class limits and the cumulative frequencies is plotted. This allows the user to determine how many or what percent of the values are equal to or more than a selected amount.

Line charts are used to portray a series of data over a period of time. Bar charts are also used to show

long-term trends. Two-directional bar charts are used to show changes from one period to another. Pie charts can be used to show percentages of a total.

Measures of Location

A measure of Central Tendency is a single value used to represent a data set. The arithmetic mean is the most often used measure of central tendency. To compute the arithmetic mean, sum the observations and divide by the total number of observations.

The geometric mean is the nth root of the product of n values. It is often used to measure the rate of growth in a series of data.

The median is the middle value in a set of values ordered from the smallest to the largest. Fifty percent of the observations are greater than the median and 50% are less than the median.

The mode is the value that occurs most often in a set of data. The mode can be determined for all levels of data and a set of data can have more than one mode.

Skewness is the lack of symmetry in data. If data is not skewed the mean, median and mode are equal. Half of the values are above the mean, median and mode and half are below. If the long tail of the distribution is to the right the data is positively skewed. If it is to the left the data is negatively skewed.

For positively skewed data the mode is the highest point of the distribution and the mean is the largest of the three averages. For negatively skewed data the mode is the highest point of the distribution and the mean is the smallest of the three averages.

Measures of Dispersion

A measure of dispersion is used to evaluate a measure of central tendency. A larger dispersion indicates that the data are more widely spread out from the mean.

The range is the difference between the highest and the lowest value.

The mean deviation is the mean of the absolute differences between each observation and the mean.

The variance is the mean of the squared deviations between the mean and each observation. Properties of the variance are:

- All values are used in the calculation.
- It is not influenced by large or small values.

The standard deviation is the square root of the variance. Properties of the standard deviation are:

- The value is in the same units as the original data.
- Chebyshev=s theorem states that regardless of the shape of the distribution at least $1 - 1/k^2$ of the observations will be within k units of the mean.

The quartile deviation is the difference between the first and third quartile divided by 2. The distance between the first quartile and the third quartile is called the interquartile range.
The empirical or normal rule states that:

- Mean plus or minus 1 Std Deviation encompasses about 68% of the values.

- Mean plus or minus 2 Std Deviations encompasses about 95% of the values.
- Mean plus or minus 3 Std Deviations encompasses about 99.7% of the values.

The coefficient of variation is a measure of relative dispersion.

The coefficient of variation should be used when there is a wide difference in the magnitude of the means being computed and when the distributions being compared are in different units.

The coefficient of skewness is a measure of the lack of symmetry in a distribution. If the long tail of the distribution is to the right the distribution is positively skewed. If the long tail of the distribution is to the left the distribution is negatively skewed.

A Survey of Probability Concepts

A probability is a value between 0 and 1 that represents the likelihood a particular event will occur. An experiment is the observation of some activity or the act of taking some measurement. An experiment is the observation of some activity or the act of taking some measurement. An outcome is a particular result of an experiment. An event is the collection of one or more outcomes of an experiment.

There are three definitions of probability:

1. the classical definition is appropriate when there are n equally likely outcomes to an experiment,
2. the relative frequency definition occurs when the number of times an event occurs is divided by the total number of observations,
3. a subjective probability is based upon whatever information is available. It is applied only when there is not enough information for another method to be used.

Two events are mutually exclusive if the occurrence of one event precludes the occurrence of another event. Events are independent if the occurrence of one event does not affect the occurrence of another event.

The rules of addition are used to combine events. The special rule of addition is used to combine events that are mutually exclusive. The general rule of addition is used to combine events that are not mutually exclusive.

The complement rule is used to determine the probability of an event happening by subtracting the probability of an event not happening from 1.

The rules of multiplication are also used to combine events. Events are independent if the occurrence of one event has no affect on the occurrence of another event. The special rule of multiplication is used to combine events that are independent. The general rule of multiplication is used to combine events that are not independent.

A joint probability is the likelihood that two or more events will happen at the same time. A conditional probability is the likelihood that an event will happen, given that another event has already happened.

Bayes theorem is a method of revising a probability, given that additional information is obtained.

Discrete Probability Distributions

A random variable is a numerical value determined by the outcome of and experiment. A probability distribution is a listing of all the outcomes of an experiment and the probability associated with each outcome A discrete probability distribution can assume only certain values. The main features are:

1. The sum of the possible outcomes is 1.00.
2. The probability of a particular outcome is between 0 and 1.
3. The outcomes are mutually exclusive.

A continuous distribution can assume an infinite number of values within a specified range.

The mean and standard deviation of a probability distribution are computed as follows:

The mean equals the weighted average of the expected values
The standard deviation equals the square root of the variance and measures the variation around the mean.

Each binomial distribution has the following characteristics: 1) each outcome is classified into one of two mutually exclusive categories, the probability of a success remains the same from trial to trial, each trial is independent, the distribution results from a count of the number of successes in a fixed number of trials.

The Poisson distribution has the following characteristics: each outcome is classified into one of two mutually exclusive alternatives, the probability of success remains the same from trial to trial, each trial is independent, the distribution results from a count of the number of successes in a fixed number of trials.

The Normal Probability Distribution

The normal distribution is a continuous probability distribution with the following major characteristics.

1. It is bell shaped and the mean, median, and mode are equal.
2. It is symmetrical.
3. It is asymptotic, meaning the curve approaches but never touches the x-axis.
4. It is completely described by the mean and the standard deviation.

There is a family of normal distributions. Each time the mean or standard deviation changes a new distribution is created.

The standard normal distribution is a particular normal distribution. It has a mean of 0 and a standard deviation of 1. By standardizing a normal distribution, we can report the distance from the mean in units of the standard deviation. The normal distribution can be used to approximate a binomial distribution under certain conditions.

Sampling Methods and Sampling Distributions

There are a number of reasons for sampling a population. Often testing is destructive and the sample item is destroyed in the process, it is often not possible to check or locate all members of the population, the results of a sample may adequately estimate the population parameter, saving time and money, it may be too time-consuming to contact all members of the population.

There are two types of samples: probability and non probability. In a probability sample all members of the population have a chance of being selected. There are several probability sampling methods. In a simple random sample all members have the same probability of being selected. In a systematic sample a random starting point is selected and then every kth item is selected. In a stratified sample the population

is divided into several groups, or strata, and then a sample is selected from each stratum. In cluster sampling the population is divided into primary units, and then samples are drawn from the primary units.

In nonprobability sampling, inclusion in the sample is based on the judgement of the person conducting the sample. Nonprobability samples may lead to biased results.

The difference between the population parameter and the sample statistic is called the sampling error. The sampling distribution of the means is a probability distribution showing all possible sample means and their probability of occurrence. For a given sample size the mean of all possible means selected from the population is exactly equal to the population mean. There is less dispersion in the distribution of the sample means than there is in the population. The variation in the sampling distribution is called the standard error of the mean.

The central limit theorem states that if the population is normal the sampling distribution of the sample means is also normal, further is the population is not normal the sampling distribution of the means approaches normality as the size of the sample is increased.

A point estimate is a single value used to estimate a population value. An interval estimate is a range of values within which the population parameter is expected to occur.

Sample size can be computed for both means and proportions. The factors that determine the sample size for a mean are, 1) the desired level of confidence, 2) the maximum allowable error, and 3) the variation in the population.

Tests of Hypothesis: Large Samples

The purpose of hypothesis testing is to check the validity of statements about a population parameter. The procedures used in hypothesis testing are:

1. State the null hypothesis H_0: and the alternate hypothesis H_a.
2. Select the level of significance. The values .1, .05 and .01 are the most commonly used. This is the probability of committing a type I error or rejecting a true null hypothesis.
3. Decide on the test statistic (z or t)
4. State the decision rule. This is the area where the null hypothesis is rejected.
5. Take a sample and make a decision regarding the null hypothesis.

Linear Regression and Correlation

Correlation analysis is used to portray the relationship (correlation) between 2 or more variables. Originated by Karl Pearson (1900) the coefficient of correlation measures the strength of the relationship between two sets of interval scaled or ratio scaled variables. It is often referred to as Pearson's R or as the Pearson product-moment correlation coefficient. If there is no relationship between the two sets of variables r will be 0. This demonstrates that the relationship is weak. As the coefficients approaches 1 the relationship strengthens. For a coefficient of 1 there is a perfect relationship. The strength of the degree of correlation does not depend on the direction of the correlation. Coefficients of +1 and -1 indicate perfect correlation.

Scatter diagrams with divergent points indicate little correlation, while diagrams with data points grouped tightly around the center of the data indicate strong correlation. The Coefficient of Correlation is a measure of the strength of the linear relationship between two sets of variables.

The Coefficient of Determination is computed by squaring the correlation coefficient. This is a proportion or percent of the variation in on variable associated with or accounted for by the variation in another variable. A high correlation coefficient or coefficient of determination does not imply causality. The relationship could be spurious. In order to determine if a correlation coefficient is statistically significant it must be tested.

Regression analysis is a technique used to develop the equation for a straight line. The regression equation is used to estimate y based upon x. It is the mathematical equation used to define the relationship between two variables. The least squares principle determines a regression equation by minimizing the sum of the squares of the vertical distances between the actual y values and the predicted values of y.

Standard error of the estimate measures the scatter or dispersion of the observed values around the line of regression. The assumptions underlying linear regression are:

1. For each value of x, there is a group of y values and these y values are normally distributed.
2. The means of these normal distributions of y values all lie on the straight line of regression.
3. The standard deviations of these normal distributions are equal.
4. The y values are statistically independent. This means that in the selection of a sample, the y values chosen for a particular x value do not depend on the y values for any other x value.

Confidence intervals report the mean value of y for a given x. Prediction intervals report the range of values for a particular value of x.

Times Series and Forecasting

A time series is a set of data collected, which covers a period of time. There are four components to a time series:

1. Trend - the secular direction of the time series.
2. Cyclical - the component of the time series, which incorporates fluctuations above, and below the trend line.
3. Seasonal - the variation in a time series, which occurs within a year. Seasonal patterns will repeat from year to year.
4. Irregular - This variation can be subdivided into two components: (a) episodic variation which is unpredictable but can be identified, i.e., fire, floods, earthquakes, etc., (b) residual variation which is random in nature.

The long run trend equation is:

$Y= = a + bt$

Where a is the intercept, b is the slope, and t is time.

The trend equation is determined using the principle of least squares. If the trend is not linear but increases by a constant percent and a least squares regression is computed logarithms should be used.

Moving averages are used to smooth the trends in a time series. Seasonal factors can be estimated using the ratio-to-moving average method:

Time Value of Money

Future Value is the amount an investment is worth after one or more periods. Compounding is the process of accumulating interest in an investment over time to earn more interest. Interest-on-interest refers to interest earned on the reinvestment of previous interest payments. Compound interest is interest earned on both the initial principal and the interest reinvested from prior periods.

Simple interest is the interest earned only on the original principal amount invested. The future value of a single amount is equal to:

$FV = PV * (1 + r)^t$
Where:
FV is the Future Value
PV is the Present Value (Value today)

Let: $(1 + r)^t$ = FVIF (the future value interest factor)

Then: FV = PV * FVIF

The present value is the current value of future cash flows discounted at the appropriate discount rate.

$PV = FV / (1 + r)^t$

The discount rate is the rate used to calculate the present value of future cash flows. The discount rate can be determined by recognizing that there are four parts to the equation in 7 above: (1) present value, (2) future value, (3) discount rate, and (4) number of periods. If any three parts are known the fourth can be determined. Cash flow timing refers to the fact that cash flows are assumed to occur at the end of each period. An annuity is a level stream of cash flows for a fixed period of time. It is a uniform stream of equal sized payments or receipts. An annuity due is an annuity for which the cash flows occur at the beginning of the period.

Annuity due value = Ordinary annuity value * (1 + r)

A perpetuity is a level stream of cash flows that occurs forever.

PV * r = CF (Cash Flow)
PV = CF / r

The effective annual rate is the interest rate expressed as if compounding occurred once a year.

Continuous compounding

Effective annual rate = $e^q - 1$

Where: q is the quoted interest rate.

The Investment Setting

Holding period return is simply the ending value of an investment divided by the beginning value of the investment. The holding period yield may be converted to a percentage rate by subtracting 1 from the holding period return.

HPY = HPR -1

Annul Holding Period Return is given by the following formula

Annual HPR = $HPR^{1/n}$

Where: n = number of years the investment is held.

Two summary measures of return performance are the arithmetic mean (dollar weighted) return and the geometric mean (time weighted) return.

The Geometric mean return is a superior measure of the long-term mean rate of return because it indicates the compound annual rate of return based on the ending value of an investment versus its beginning value.

The Arithmetic mean return is a good measure of expected rate of return when one future year is being evaluated. The geometric mean will equal the arithmetic mean when rates are the same for all years. If returns vary the GM will be less than AM. The greater the volatility in returns the greater the difference between the two return measures. The mean historical rate of return (HPY) for portfolios is measured as the weighted average of the HPYs for the individual investments in the portfolio. The weights used are the relative beginning market values for each investment. This is referred to as the dollar-weighted or value-weighted mean rate of return.

Study Session 2, Investment Tools: Quantitative Methods, DeFusco, McLeavey, Pinto and Runkle, Quantitative Methods for Investment Analysis.

Learning Outcomes.

In your study, be sure to pay close attention to the command words. For example, there is a big difference between the words “define” and “calculate” that affects your exam preparation.

A) The Time Value of Money, Ch. 1

Be very familiar with either your calculator or compound interest tables. You will be given tables for use on the exam. I highly recommend that you use both when preparing for the exam to make sure that you determine the same solution. Selected end-of-chapter problems: 11, 15, 19, 20.

a) Calculation of a future value given a present value is called compounding; calculation of a present value given a future value is called discounting.

b) Four values possible in single sum problems—given any three, you can always determine the fourth

c) FV annuity—depends on interest rate (r), payment (cash flow = A), number of periods (n), and whether ordinary (payments at end of period) or annuity due (payments at beginning of period). If annuity due, imagine a regular annuity plus one more period of compounding using a lump sum.

d) Annuity problems—given any three values (out of four), you can always calculate the fourth.

e) Present value = Future value n periods from now discounted to present at the discount rate r. You can think of the present value of an annuity as the sum of multiple lump sum present values.

f) PV perpetuity—divide the constant payment by the discount rate. Think of the DDM with g = 0. PV = A/r

g) Unknown variable in perpetuity problems—given two values (out of three), you can always determine the third.

h) For FV, compound forward each uneven CF separately and sum together. For present value, discount back to the present each uneven CF separately and sum together.

i) Non-annual compounding—whether PV or FV, divide the interest rate by the number of compounding periods (n) and multiple the number of years by n. Remember there are 2 adjustments.

j) Stated rate is the effective rate if compounding is annual. If compounding is other than annual (i.e., quarterly), compound $1 at 1 plus the periodic rate (stated rate/4) raised to the fourth power and then subtract 1 to determine the effective rate.

k) Effective rate—same as j above.

l) Time lines are a very effective means of visualizing a time value of money problem. Use time lines when practicing all problems in order to be proficient with this technique for the exam.

B) Statistical Concepts and Market Returns, Ch. 3

Selected end-of-chapter problems: 8, 9, and 16.

a) A population includes all members of a specified group. A sample is a subset of a population

b) Parameter—is a descriptive measure of a population

c) Measurement scales can be nominal (weakest measure such as small-cap), ordinal (observations placed into separate categories such as top 10 of 100 mutual funds assigned the number 1), interval (provide both ranking and differences between values are equal such as temperatures) or ratio (strongest measure with zero point as origin and characteristics of interval such as rates of return)

d) A frequency distribution is a tabular display of data summarized into a relatively small number of intervals such as holding period returns from 0% to 10%.

e) Holding period return—is the total return including price appreciation (growth) plus yield (income). On exam, be sure to recognize both components.

f) An interval is a set of return values within which an observation falls.

g) Relative frequencies are found by dividing the absolute frequency of each return interval by the total number of observations.

h) A histogram is a graphical equivalent of a frequency distribution presented as a bar chart.

i) Central tendency measures include the population mean (population sum total/number of observations), sample mean (sample sum of total/number of observations), arithmetic mean (also called the sample mean or simple mean), geometric mean (compound average found by multiplying 1 plus the periodic return, taking the nth root of the product, and subtracting 1.

j) Geometric mean equals the arithmetic mean when there is no variability of returns. The difference increases as the variability increases. Use geometric mean for measuring past performance and the arithmetic mean for projecting returns. The arithmetic mean is an unbiased estimator whereas the geometric mean is biased.

k) quartiles divide the distribution into quarters, quintiles into fifths, deciles into tenths and percentiles into hundredths.

l) portfolio return as a weighted mean (past performance) weighs each return in investment i by the percentage of investment in i, (2) a weighted average uses the same technique but is forward looking instead of backward, (3) a range is the difference between the highest and lowest returns; a mean absolute deviation ignores the signs of eh deviations around the mean and is superior to the range, (4) a population variance is the average of the squared deviations around the population mean using the total number of observations in the population, a sample variance is the average of the squared deviations around the sample mean using the sample size minus 1; a standard deviation is the square root of variance.

m) Chebyshev's inequality gives the proportion of elements within k standard deviation s of the mean (the proportion of the observations within k standard deviations of the mean = 1 – 1/k^2). It holds for both sample and populations, and for discrete and continuous data.

n) Coefficient of variation equals sample standard deviation divided by sample mean.

o) Sharpe measure is the portfolio's mean return minus the mean return on the risk-free asset divided by the standard deviation of returns. Note that the Sharpe measure is the slope of the CML.

p) Nonsymmetrical distributions are skewed, meaning less than 50 percent of frequency distribution is on one side of the mean.

q) Skewness is a measure of how nonsymmetrical a distribution is. You will not have to calculate this measure on the exam, but you will need to know what it means. (Note that the command word for skewness is define, not calculate). Measurement equals the cubed deviation around the mean whereas the standard deviation equals the squared deviation around the mean.

r) Kurtosis—again, the command word is define. Kurtosis is a measure of peakedness of the distribution. A distribution may have positive kurtosis when more returns are clustered around the mean and more returns with larger deviations than suggested by a normal distribution.

s) For all normal distributions, kurtosis is equal to 3. Excess kurtosis equals kurtosis minus 3. A normal distribution has excess kurtosis equal to 0, and a leptokurtic distribution has excess kurtosis greater than 0. A sample excess kurtosis of 1.0 or larger would be considered unusually large.

t) Semilog scale is used for measuring past performance rather than an arithmetic scale. Semilog is

similar to using a geometric growth rate.

C) Probability Concepts, Ch. 4

(be sure to distinguish between past events and future events where probabilities are used to calculate expected values). Selected end-of-chapter problems: 1, 10,12, 15, 17.

a) A random variable is the return on a risky asset whose outcome is uncertain. An outcome is a particular value of a random variable. An event is an outcome or specified set of outcomes of a random variable. Mutually exclusive events means that only one event can occur at a time. Exhaustive events means that the events cover all possible outcomes.

b) A probability is a number between 0 and 1 that gives the chance that a stated even occurs.

c) Empirical probabilities occur when the probability of an event is estimated from data.

d) Inconsistent probabilities of an event occurring applied to two investments that are similar in every way results in an arbitrage opportunity. In other words, prices of two stocks should react differently given inconsistent (unequal) probabilities being applied to each stock of an upcoming event.

e) Unconditional probabilities occur when the probability of an event is not dependent on some other event. This is called a marginal probability.

f) Joint probabilities describe the probability of two events occurring simultaneously.

g) Multiplication rule says that the joint probability of A and B equals the marginal probability of A occurring given that B occurs times the probability of B occurring. $P(AB) = P(A|B)P(B)$.

h) Addition rule says that the probability that A or B occurs, or both occur, is equal to the probability that A occurs, plus the probability that B occurs, minus the probability that both A and B occur. In other words, subtract the overlap between A and B from the total probability of either occurring.

i) Dependent events are when the probability of one event occurring is related to the probability of another event occurring. Independent events are unrelated.

j) The joint probability of independent events occurring is found by multiplying the individual probabilities of the two events together (pg. 182).

k) The total probability rule explains the unconditional probability of the event in terms of probabilities conditional on various scenarios. The scenarios must be mutually exclusive and exhaustive (pg. 185).

l) Expected value of a random variable is the probability-weighted average of the possible outcomes of the random variable. Variance of a random variable is the expected value of squared deviations from the random variable's expected value and standard deviation is the square root of variance (pg. 188).

m) Conditional expectation involves adjustments based on new information or events

n) Total probability rule for expected value is in terms of conditional expected values: E(X) = E(X|S1)P(S1) + E(X|S2)P(S2) + + E(X|Sn)P(Sn) (pg 190).

o) Covariance between two random variables is the probability-weighted average of the cross-product of each random variable's deviation from it own expected value (pg 197).

p) Relationships: correlation (ij) equals covariance (ij) divided by standard deviation (i) time standard deviation (j). Note that as covariance increases, correlation increases *ceteris paribus*, and as standard deviations of either increase, correlation decreases *ceteris paribus* (pg. 200).

q) Expected return on a portfolio is the weighted average return; variance is the weighted average of covariances (pg 198).

r) Covariance is the weighted average of the cross-products of each random variable's deviation from its own expected value. (pg 199).

s) Bayes' formula is a methodology for adjusting expectations given new information (pg 206 example 4-13).

t) Multiplication rule is n!, which is read n factorial. Thus, the number of ways k things can occur = n1 x n2 x x nk.

u) Factorial is n!, combination is nCr, or the number of ways of selecting r objects from a total of n objects, where the order in which the r objects is listed does not matter. nCr=n!/(n-r)! x r!. Permutation is nPr = n!/(n-r)!, which is the number of ways to select r objects from a total of *n* objects, where the order is which the r objects is listed does matter (pg 210).

v) Problems for which different counting methods are appropriate depends on whether order matters (Pg 210).

w) When order does not matter use combination; when order does matter, use permutation.

D) Common Probability Distributions, Ch. 5

Selected end-of-chapter problems: 12-15.

a) Probability distribution specifics the probabilities of the possible outcomes of a random variable.

b) Discrete random variables can take on at most a countable number of possible values. Continuous random variables can take on an infinite number of possible values.

c) Range is defined from the lowest to the highest possible value.

d) Probability function specifies the probability that the random variable takes on a specific value (PX = x) is the probability that a random variable X takes on the value x.

e) Key properties of a probability function are that the probabilities must lie between 0 and 1, and the sum of all probabilities must equal 1 (pg 228).

f) Cumulative distribution function gives the; probability that a random variable X is less than or equal to a particular value x, P(X is less than or equal to x).

g) A probability density function is a distribution that completely describes a continuous variable. It describes the probability that a random variable takes on a specific value P(X = x) (pg 228).

h) Discrete uniform random variable is where the distribution has a finite number of specified outcomes, and each outcome is equally likely. (pg 228).

i) Binomial probability distribution has one of two possible values: success or failure.

j) Expected value equals the probability-weighted average of the possible ending values. The variance equals the weighted average of the squared deviations from the mean (pg 239, example 5-6).

k) A binomial tree is a graphical description of possible successes and failures (pg 238).

l) Continuous uniform distributions describe equal probabilities of a continuous random variable (pg 241).

m) A normal distribution is completed defined by two parameters: mean and standard deviation. It is symmetrical and a linear combination of two or more normal random variables is also normally distributed (pg 244-245).

n) Confidence intervals depend on the mean and standard deviation (pg 246). Be sure to know the following confidence intervals: 68 percent (+/- 1 standard deviation), 95 percent (+/- 2 standard deviations), and 98 percent (+/- 3 standard deviations) (pg 246).

o) Standard normal distribution has 0 mean and standard deviation of 1. Subtracting the mean of X from X and dividing the result by the standard deviation gives a Z variable (pg 248).

p) Probabilities for the standard normal distribution are found in Appendix A (partial probabilities on pg 249).

q) Univariate describes a single random variable and is described by two parameters: mean and variance. Multivariate specifies the probabilities for a group of related random variables and is described by three parameters: mean, variance, and covariance (pg 245).

r) Correlation in multivariate distributions drives covariance (pg pg 245).

s) Shortfall risk is the risk that portfolio value will fall below some minimum acceptable level over some time horizon (pg 253).

t) Roy's safety-first rule states that the optimal portfolio minimizes the probability that portfolio return falls below the threshold level (pg 253).

u) A return is follows a lognormal distribution if it natural log is normally distributed. Remember "the log is normal" (pg 257).

v) Discretely returns have finite compounding periods such as quarterly. Continuously compounded returns associated with a holding period is the natural log of 1 plus the holding period return, or, equivalently, the natural log of the ending price over the beginning price (the price relative) (pg 258).

w) Continuous compounded return is ln S(t+1)/St = ln (1+Rt,t+1) (pg 258).

x) Monte Carlo simulation involves the use of a computer to represent the operation of a complex financial system based on random samples from a probability distribution (pg 261).

Study Session 3, Investment Tools: Quantitative Methods, DeFusco, et al.

Learning Outcomes

A) Sampling and Estimation, Ch. 6
Selected end-of-chapter problems: 3, 9, 10, 14.

a) Random sampling is obtained such that each element of the population has an equal probability of selection.

b) Sampling error is the difference between the observed value of a statistic and the quantity it is intended to estimate.

c) Sampling distribution is the distribution of all the distinct possible values that the statistic can assume when computed from samples of the same size randomly drawn from the same population.

d) Stratified means that the population is subdivided into subpopulations (strata) based on one or more classifications. Simple random means that the sample comes from the whole population (pg 281).

e) Time-series is a sequence of returns collected at discrete and equally spaced intervals of time. Cross-sectional data come from groups at a specific point of time.

f) Central limit theorem states that the probability distribution having mean and finite variance, the sampling distribution of the sample mean computed from samples of size n from this population will be approximately normal when the sample size is large. Its importance is by placing importance on sample size whatever the distribution of the population (pg 286).

g) Standard error of sample mean equals the sample standard deviation, s, divided by the square root of the sample size, n. Use the standard error to make probability statements about the population mean via confidence intervals.

h) Point estimate is a single number that estimates the unknown population parameter. A confidence interval gives a range of possible values for the parameter at some specified level of probability.

i) Desirable properties of an estimate are unbiasedness (estimator's expected value equals the parameter it is intended to estimate), efficiency (if no other unbiased estimator of the same parameter has a sampling distribution with small variance), and consistency (the probability of accurate estimates increases as sample size increases).

j) A confidence interval is an interval for which you can assert with a given probability 1-alpha, called the degree of confidence, that it will contain the parameter it is intended to estimate.

k) Student's t is a symmetrical probability distribution defined by a single parameter known as degrees of freedom (df). Use this distribution when the sample size is small and the population variance is unknown.

l) Confidence interval for a population mean is given by: sample mean +/- t-value (alpha divided by 2) times sample standard deviation divided by square root of n (pg 296).

m) Sample size depends on whether distribution is normal and whether variance is known. If distribution is normal and variance is known, use z even if samples size is small. If distribution is normal and variance is unknown, use t is sample size is small. If sample size is large, always use z whether distribution is normal or not and whether variance is known or not (pg 298).

n) Data mining occurs from extensive searching through databases for pattern or trading rules (repeated drilling the data for results, pg 298).

o) Selection bias occurs when data availability leads to certain assets being excluded from the analysis (pg 299).

B) Hypothesis Testing, Ch. 7

Selected end-of-chapter problems: 1-5.

a) Steps include: (1) state hypothesis, (2) identify test statistic and its probability distribution, (3) specify level of significance, (4) state decision rule, (5) collect data and perform calculations, (6) make statistical decision, (7) make economic decision pg 317).

b) Null hypothesis is the hypothesis to be tested. It is stated as equality. Alternative is hypothesis accepted when null is rejected. It is stated as complement to null (pg 318).

c) For a one-tail-test, the alternative hypothesis gives rejection in only one tail of distribution. For a two-tail test, the alternative gives rejection in two tails of the distribution.

d) A test statistic is a quantity calculated on the basis of a sample, whose value is the basis for deciding whether to reject or not reject eh null hypothesis.

e) Significance level is the probability of a Type I error (rejecting the null when it is true) in testing a hypothesis.

f) Type I error is rejecting the null when it is true and a Type II error is not rejecting null when it is false (Table 7-1, pg 321).

g) Power of the test is the probability of correctly rejecting the null (pg 321).

h) Decision rule state when to accept or reject the null hypothesis. When rejecting the null, the result is statistically significant (pg 322).

i) Confidence intervals represent an alternative method of tests of significance. When the hypothesized value of the population parameter under the null is outside the corresponding confidence interval, the null hypothesis is rejected (pg 324).

j) Statistical decision is based on empirical data contained in the test. An economic decision is

based on the statistical decision, but also all economic issues pertinent to the decision (pg 325).

k) p-value is the smallest level of significance at which the null hypothesis can be rejected (325).

l) Test statistic is t when population variance is unknown (pg 327), but z when population variance is known (pg 330).

m) z-test is appropriate for mean testing when the sample size is large since the Central Limit Theorem says that the distribution of the mean is normal when n is large.

n) The test statistic for equality of two population means from independent samples is a t-statistic. (pg 334).

o) Test statistic for a paired comparison (samples are not independent) test is a t-statistic (pg 338). Null hypothesis: mean1difference = mean2difference where mean1difference equals difference between means of paired samples (Table 7-5. pg 338).

p) Test statistic is a chi-square (pg 341). Null hypothesis is that the hypothesized value equals some specific value; the alternative hypothesis is that they are unequal. If the test statistic (chi-square) falls outside the acceptance ranges, the null hypothesis is rejected, and vice versa (pg 341).

q) The test statistic is an F-test, which is the ratio of sample variances (pg 343). The null hypothesis is that the two variances are equal. The alternative is that they are not equal. Reject the null if the test statistic falls outside the acceptance region, and vice versa (pg 344).

r) Parametric tests assume that the test statistic used in the hypothesis is drawn from a population that has a certain distribution, such as normality (pg 345).

C) Correlation and Regression, Ch. 8

Selected end-of-chapter problems: 5, 11, 17.

a) Scatter plot is a graph that shows the relationship between the observations for two data series in two dimensions (pg363).

b) Covariance is a measure of linear relationship between two variables. It has no boundaries (pg

c) Correlation coefficient is a measure of linear relationship between two variables. It is bound by –1 and +1 (pg 365).

d) If the correlation coefficient equals 1.0 (-1.0), the relationship between two variables is perfectly positive (negative). That is, by knowing the independent variable, you will know exactly the dependent variable.

e) The null hypothesis is that r = 0. The alternative is r does not equal 0. If the test statistic, t, falls outside the acceptance region, reject the null hypothesis, and vice versa (pg 377).

f) Outliers are small numbers of observation at either extreme of a sample. Outliers weaken the correlation coefficient (pg 368).

g) Spurious correlation is not based on any theoretical relationship (pg 370).

h) Dependent variable (left hand side of equality sign) derives its value from the independent variable (right hand side of equality sign). The independent variable explains changes in the dependent variable (pg 379).

i) In a two-variable regression, the slope measures the degree of the relationship. A slope greater than 1.0 means a small change in the independent variable results in a larger change in the dependent variable, and vice versa. The intercept gives the value of the dependent variable when the independent variable equals zero (pg 380).

j) Assumptions of linear regression are: (1) linear relationship, (2) independent variable is not random—this is a simplifying assumption, but not essential, (3) expected value of error term is zero, (4) variance of error term is constant across all observations of independent variable, (5) error terms are uncorrelated, (6) error term is normally distributed. All of these assumptions are made for convenience, but we have sophisticated techniques for adjusting the regression technique (pg 383).

k) Standard error of the estimate measures the uncertainty of the regression equation when estimating the dependent variable. The larger the error term, the larger the SSE, and vice versa (pg 386).

l) Coefficient of determination measures the fraction of the total variance in the dependent variable that is explained by the independent variable (pg 388).

m) Confidence intervals are intervals that we believe include the true parameter value being tested and depend on the SEE (pg 391).

n) The test statistic is a t since the regression equation is estimating more than one variable. In the case of a simple regression, the t-statistic has n-2 df since we are estimating a slope and an intercept. For each additional variable being estimated, we lose a df (391).

o) Regression coefficients allow estimation of the dependent variable given values of the independent variable(s) (pg 380).

p) Predicted values come from the regression equation that depends on the accuracy of the regression coefficients (pg 399).

q) The confidence interval for the estimated (or predicted) value of the dependent variable depends on the standard deviation of the forecast error, which depends on the SEE, number of observations, value of independent variable, estimated mean of the independent variable, and the variance of the independent variable (pg 400-401).

r) ANOVA is a statistical procedure for analyzing the total variability of a set of data into components that can be attributed to different sources. It is useful in determining whether the independent variables can explain variation in the dependent variable (pg 396).

s) In a simple regression, the F-statistic duplicates information contained in the t-statistic. In multiple regression, the two are not duplicative. The F-statistic tests whether all the slope coefficients in the regression equation equal zero (pg 396).

t) Limitations of regression analysis include: (1) relations can change over time, (2) widespread knowledge allows other analysts to take advantage of analysis, (3) many assumptions stated in j

above are violated in practice (pg 403).